

**1. What does Bayes' Theorem describe?**

- A) A method for clustering data
- B) The probability of an event based on prior knowledge ✓
- C) The relationship between dependent variables
- D) A function for dimensionality reduction

**2. What is the primary assumption of the Naïve Bayes classifier?**

- A) All features are dependent
- B) All features are independent given the class ✓
- C) Data must be normally distributed
- D) Only categorical features are allowed

**3. The Gibbs Algorithm is mainly used for:**

- A) Optimizing machine learning models
- B) Bayesian probability estimation ✓
- C) Image classification
- D) Time series forecasting

**4. What is Maximum Likelihood Estimation (MLE) used for?**

- A) Minimizing feature space
- B) Finding parameters that maximize the probability of observed data ✓
- C) Reducing variance in predictions
- D) Clustering large datasets

**5. The Minimum Description Length (MDL) principle aims to:**

- A) Find the most complex explanation for a dataset
- B) Find the simplest explanation for a dataset ✓
- C) Increase the number of features
- D) Reduce computation speed

**6. In Naïve Bayes, the probability of an event given evidence is known as:**

- A) Prior probability
- B) Posterior probability ✓
- C) Likelihood
- D) Marginal probability

**7. What is the role of the prior probability in Bayes' Theorem?**

- A) It represents initial beliefs before new evidence is considered ✓
- B) It is the probability of the evidence
- C) It is the final classification output
- D) It is always equal to 1

**8. What type of learning is K-Nearest Neighbors (KNN)?**

- A) Supervised Learning ☒
- B) Unsupervised Learning
- C) Reinforcement Learning
- D) Semi-supervised Learning

**9. How does KNN classify a new data point?**

- A) By assigning the majority class of its nearest neighbors ☒
- B) By learning parameters during training
- C) By computing probability distributions
- D) By reducing dataset size

**10. Which distance metric is commonly used in KNN?**

- A) Euclidean Distance ☒
- B) Hamming Distance
- C) Jaccard Similarity
- D) Cosine Similarity

**11. What is the main disadvantage of KNN?**

- A) High training time
- B) High computation during inference ☒
- C) Poor accuracy
- D) Requires large amounts of labeled data

**12. In which industry is Naïve Bayes frequently used?**

- A) Medical diagnosis
- B) Spam filtering ☒
- C) Self-driving cars
- D) Image processing

**13. The key difference between supervised and unsupervised learning is:**

- A) Supervised learning does not require labels
- B) Unsupervised learning uses labeled data
- C) Supervised learning requires labeled data ☒
- D) Unsupervised learning uses target variables

**14. Which of the following is an example of unsupervised learning?**

- A) Decision Trees
- B) K-Means Clustering ☒
- C) Naïve Bayes
- D) Logistic Regression

**15. What is an advantage of the Naïve Bayes classifier?**

- A) Handles missing data well
- B) Works well with small datasets ✓
- C) Provides 100% accuracy
- D) Requires deep neural networks

**16. Which step is NOT involved in Bayes' Theorem computation?**

- A) Calculating the prior probability
- B) Calculating the likelihood
- C) Calculating the gradient descent ✓
- D) Calculating the posterior probability

**17. Why is Naïve Bayes called "naïve"?**

- A) It does not learn from training data
- B) It assumes feature independence ✓
- C) It uses deep learning
- D) It only works for small datasets

**18. What is a key characteristic of Instance-Based Learning?**

- A) It requires explicit rule-based training
- B) It memorizes training examples for making predictions ✓
- C) It is used only for regression problems
- D) It requires a pre-trained neural network

**19. What is a real-world application of KNN?**

- A) Face recognition
- B) Fraud detection ✓
- C) Stock market prediction
- D) Game playing

**20. Which of the following is NOT an application of Machine Learning?**

- A) Speech recognition
- B) Image classification
- C) Cooking food ✓
- D) Spam email detection

**21. What is the main goal of supervised learning?**

- A) To find hidden patterns in data
- B) To learn a mapping from inputs to outputs using labeled data ✓
- C) To cluster similar data points
- D) To reduce the dimensionality of data

---

**22. In which of the following methods does the algorithm store all training instances and classify new data based on similarity?**

- A) Decision Trees
  - B) K-Nearest Neighbors (KNN) ☒
  - C) Support Vector Machines (SVM)
  - D) Random Forest
- 

**23. What is the primary limitation of KNN?**

- A) It is difficult to implement
  - B) It requires large amounts of labeled data
  - C) It has high computational cost for large datasets ☒
  - D) It cannot handle non-numeric data
- 

**24. Which of the following is a distance-based method?**

- A) Support Vector Machines
  - B) K-Means Clustering
  - C) K-Nearest Neighbors ☒
  - D) Neural Networks
- 

**25. Which of the following statements about Decision Trees is true?**

- A) They perform well with non-linear relationships ☒
  - B) They are sensitive to missing data
  - C) They do not require labeled data
  - D) They use support vectors for classification
- 

**26. Which of the following is a key advantage of Decision Trees?**

- A) High accuracy for large datasets
  - B) Easy interpretability ☒
  - C) Requires no hyperparameter tuning
  - D) Works well with high-dimensional data
-

**27. Which of the following methods is based on finding a hyperplane that maximizes the margin between two classes?**

- A) KNN
  - B) Decision Trees
  - C) Support Vector Machines (SVM) ☒
  - D) K-Means Clustering
- 

**28. What is the kernel trick used for in SVM?**

- A) To convert categorical features into numerical values
  - B) To transform non-linearly separable data into higher-dimensional space ☒
  - C) To reduce the number of training examples
  - D) To optimize hyperparameters
- 

**29. Which of the following is NOT a kernel function in SVM?**

- A) Linear Kernel
  - B) Polynomial Kernel
  - C) Gaussian Kernel
  - D) K-Means Kernel ☒
- 

**30. In which type of learning is clustering used?**

- A) Supervised Learning
  - B) Unsupervised Learning ☒
  - C) Reinforcement Learning
  - D) Semi-supervised Learning
- 

**31. What is the goal of clustering in machine learning?**

- A) To classify data based on prior labels
  - B) To partition data into meaningful groups based on similarity ☒
  - C) To maximize classification accuracy
  - D) To remove outliers
-

**32. Which of the following is NOT a clustering algorithm?**

- A) K-Means
  - B) Hierarchical Clustering
  - C) Decision Trees ☒
  - D) DBSCAN
- 

**33. What is the main drawback of K-Means clustering?**

- A) It is very slow
  - B) It is sensitive to the choice of initial cluster centers ☒
  - C) It requires labeled data
  - D) It can only cluster numerical data
- 

**34. In K-Means clustering, what does the 'K' represent?**

- A) Number of iterations
  - B) Number of clusters ☒
  - C) Number of dimensions in data
  - D) Number of training examples
- 

**35. What is dimensionality reduction used for in machine learning?**

- A) To increase the number of features
  - B) To improve model interpretability and reduce computation ☒
  - C) To generate more training data
  - D) To remove important information from data
- 

**36. Which of the following is NOT a dimensionality reduction technique?**

- A) Principal Component Analysis (PCA)
  - B) t-SNE
  - C) K-Means Clustering ☒
  - D) Singular Value Decomposition (SVD)
-

**37. What does PCA aim to do?**

- A) Find the best linear separation between two classes
  - B) Transform data into a new set of orthogonal variables (principal components) ☒
  - C) Reduce data size by removing missing values
  - D) Increase the number of features
- 

**38. What is the main limitation of PCA?**

- A) It works only for classification problems
  - B) It assumes linear relationships in the data ☒
  - C) It requires a very large dataset
  - D) It cannot be used for dimensionality reduction
- 

**39. Which of the following is true about kernel PCA?**

- A) It is an extension of PCA that allows for non-linear transformations ☒
  - B) It requires labeled data
  - C) It is a supervised learning method
  - D) It is not useful for image processing
- 

**40. In which scenario would you use Kernel PCA instead of standard PCA?**

- A) When data has a non-linear structure ☒
- B) When there are too many missing values
- C) When the dataset is very large
- D) When labels are available for classification

**41. What is the primary purpose of evaluating machine learning algorithms?**

- A) To increase the number of features in the model
  - B) To compare different models and select the best one ☒
  - C) To reduce the amount of training data
  - D) To convert data into numerical form
- 

**42. Which metric is commonly used for evaluating classification models?**

- A) Mean Squared Error (MSE)
- B) Accuracy ☒
- C) R-squared
- D) Adjusted R-squared

---

**43. Which of the following is NOT a method for model selection?**

- A) Cross-validation
  - B) Hyperparameter tuning
  - C) Overfitting ☒
  - D) Grid search
- 

**44. What is cross-validation used for in machine learning?**

- A) To test different machine learning models on different datasets
  - B) To split the dataset into multiple parts for more reliable model evaluation ☒
  - C) To create new features from existing data
  - D) To apply deep learning techniques
- 

**45. Which of the following statements about ensemble learning is true?**

- A) It combines multiple weak models to create a strong model ☒
  - B) It always uses decision trees
  - C) It works only for supervised learning
  - D) It cannot be used for deep learning
- 

**46. What is the key difference between bagging and boosting?**

- A) Bagging reduces variance, while boosting reduces bias ☒
  - B) Boosting increases variance, while bagging reduces it
  - C) Both methods increase bias
  - D) Boosting is faster than bagging
- 

**47. Which ensemble method creates multiple decision trees and aggregates their predictions?**

- A) K-Nearest Neighbors
  - B) Random Forest ☒
  - C) Logistic Regression
  - D) Principal Component Analysis
-



**48. What is the role of weak learners in boosting?**

- A) They overfit the data
  - B) They are combined sequentially to improve model performance ✓
  - C) They work independently
  - D) They ignore misclassified instances
- 

**49. Which of the following is an example of a boosting algorithm?**

- A) Random Forest
  - B) AdaBoost ✓
  - C) K-Means
  - D) PCA
- 

**50. What type of data is typically used in time-series modeling?**

- A) Unstructured data
  - B) Sequential data with timestamps ✓
  - C) Randomized categorical data
  - D) High-dimensional image data
- 

**51. Which method is commonly used for time-series forecasting?**

- A) Decision Trees
  - B) ARIMA ✓
  - C) Naïve Bayes
  - D) Support Vector Machines
- 

**52. Which of the following deep learning models is commonly used for sequence modeling?**

- A) Convolutional Neural Networks (CNNs)
  - B) Recurrent Neural Networks (RNNs) ✓
  - C) Random Forest
  - D) K-Means Clustering
-

**53. What is a Deep Boltzmann Machine (DBM) primarily used for?**

- A) Supervised learning tasks
  - B) Feature learning and representation learning ☒
  - C) Clustering
  - D) Decision Trees
- 

**54. What is the main purpose of autoencoders?**

- A) To classify data
  - B) To reconstruct input data by learning compressed representations ☒
  - C) To cluster data
  - D) To generate decision trees
- 

**55. Which component is used in autoencoders to compress data into a lower-dimensional form?**

- A) Decoder
  - B) Encoder ☒
  - C) Activation Function
  - D) Loss Function
- 

**56. How do deep generative models differ from traditional machine learning models?**

- A) They rely solely on labeled data
  - B) They generate new data similar to the training data ☒
  - C) They only perform classification tasks
  - D) They require supervised learning
- 

**57. Which of the following is a deep generative model?**

- A) Logistic Regression
  - B) Variational Autoencoder (VAE) ☒
  - C) Random Forest
  - D) Decision Tree
-

**58. What is the primary application of deep networks in Natural Language Processing (NLP)?**

- A) Image classification
  - B) Sentiment analysis and machine translation ☒
  - C) Clustering
  - D) Time-series forecasting
- 

**59. In deep learning, which of the following is an application of Convolutional Neural Networks (CNNs)?**

- A) Time-series forecasting
  - B) Image recognition and object detection ☒
  - C) Sentiment analysis
  - D) Dimensionality reduction
- 

**60. What is a major advantage of using deep learning in healthcare applications?**

- A) It reduces the need for domain expertise
- B) It automatically learns features from medical images and data ☒
- C) It replaces all doctors
- D) It only works with structured data